Introducción a la creación de páginas web

Pere Barnola Augé

XP08/93133/01327



Pere Barnola Augé

Primera edición: septiembre 2008

Avda. Tibidabo, 39-43, 08035 Barcelona
Realización editorial: Eureca Media, SL

Introducción

Breve historia del World Wide Web y la hipermedia

Comenzaremos con el concepto básico: **hipermedia**.

La hipermedia es el conjunto de procedimientos o métodos que agrupan texto, vídeo, audio u otros tipos de información con referencias cruzadas entre los contenidos, con la principal característica de que el usuario puede interaccionar con ellas.

Ejemplo

Imaginad un documento en que se habla del Everest, y que en un punto de este documento aparece el nombre de "Edmund Hillary", que fue el primer escalador que alcanzó su cima. En un documento convencional, lo más probable es que tuviéramos un asterisco, o un número que hiciera referencia a una anotación al pie de página donde se explicarían brevemente las proezas de dicha persona. Pues bien, la hipermedia nos permitiría que el nombre fuera un enlace, y que, cuando el usuario hiciera clic (interaccionara), accediéramos a otro documento con toda la información de Edmund Hillary, con la opción de encontrar más referencias en ese documento, entre ellas, una al documento del Everest, o incluso imágenes, vídeos..., con lo que se podría construir una gran red de contenidos con referencias entrecruzadas.

El ejemplo más claro de hipermedia es el **World Wide Web**, que no es nada más que un sistema de documentos de hipermedia enlazados y accesibles por Internet.

Actualmente, con un navegador web (Explorer, Firefox, etc.), un usuario visualiza páginas web con diferentes tipos de contenidos (texto, imágenes, vídeos) y puede navegar por ellos utilizando los hipervínculos.

Un poco de historia

La idea inicial de web se remonta a 1945, cuando el director de la Oficina de Desarrollo e Investigación Científica (Estados Unidos), el doctor Vannevar Bush, escribió el artículo "As We May Think", en el cual expresaba su preocupación por la gran cantidad de información que existía y se estaba generando, y los sistemas poco eficientes para ordenarla y encontrarla. Así, basándose en la tecnología de aquellos tiempos, describió el Memex, que imaginaba como un suplemento a su memoria, y que permitiría a cualquier persona guardar su información en microfilms para poder consultarla posteriormente de forma rápida, y lo que es más importante, creó vínculos entre unos documentos y los otros, de manera que durante la lectura de un documento se recordara al lector cuáles eran los que contenían información relacionada. Era una visión de lo que pasaría cuarenta y cinco años más adelante.

El término *hipertexto* fue utilizado por primera vez por Ted Nloson en 1965, quien ideó un modelo para la interconexión de documentos electrónicos. Antes, sin embargo, habían surgido otros proyectos como el NLS (oNLine System), de Douglas Engelbart, un entorno de trabajo por ordenador, con teclado, pantalla, ratón e impresora, con posibilidad de teleconferencia y correo electrónico mediante una red de ordenadores para la comunicación rápida entre los profesionales.

El World Wide Web fue inventado en 1989 por un informático del CERN (Organización Europea de Investigación Nuclear) llamado Tim Berners-Lee. Era un sistema de hipertex-

to para compartir información basado en Internet, concebido originalmente para servir como herramienta de comunicación entre los científicos nucleares del CERN. A finales de 1990, el primer navegador de la historia, el WorldWideWeb, ya tenía forma.

A principios de 1993 había en torno a cincuenta servidores (ordenadores que alojaban webs). Había dos tipos de *browsers* o navegadores para cualquier plataforma, pero muy limitados y muy poco atractivos. En febrero, se lanzó la primera versión alfa del navegador Mosaic for X, desarrollado en el NCSA (National Center for Supercomputing Applications). Funcionaba en X Windows, que era una plataforma popular entre la comunidad científica. En abril, el tráfico del WWW era el 0,1% del total de Internet. El CERN declaraba el WWW como tecnología de acceso gratuito. En septiembre, ya había versiones de mosaico para PC y Macintosh. El tráfico llegaba al 1% de todo el tráfico de Internet y había más de 500 servidores. Es el comienzo del crecimiento explosivo de la web. A finales de 1994, ya había más de 10.000 servidores y 10 millones de usuarios. Y en 1997, más de 650.000 servidores.

Hoy en día, la web es algo cotidiano para una gran parte de los más de 600 millones de usuarios de Internet que hay en todo el mundo. Sus utilidades son variadas, y su impacto en la economía mundial es más que apreciable. No sólo existen documentos de texto, sino que también hay imágenes, vídeos, música; además, se puede comprar cualquier producto de todo el mundo, hacer reservas, etc.

Cada uno de nosotros puede ir tejiendo una parte de esta gran red...

Los estándares

Cuando el volumen en Internet empieza a crecer, se hace necesario una estandarización que vuelva a situar el intercambio de información accesible como eje principal, y establezca las bases del siguiente paso en la evolución de la web; una web basada en un código semántico que aporte más significado a la información, para facilitar su localización y recuperación.

Los estándares que harán que eso sea posible son desarrollados por un consorcio integrado por diferentes organizaciones de diferentes países, conocido como el**World Wide Web Consortium (W3C)**.

Este consorcio fue creado a mediados de los años noventa por Tim Berners-Lee.

Su principal propósito es que la web evolucione y desarrolle todo su potencial de forma ordenada, y sin perder de vista su concepto inicial de intercambio de información para que sea accesible, interoperable e independiente de tecnologías propietarias. Los estándares se desarrollan en un proceso abierto en que cualquier empresa puede participar, pero nunca ser propietaria.

Finalmente, depende de nosotros respetar el principio de universalidad o no. Cada vez que alguien crea una página de manera que no es accesible para todo el mundo, o que sólo es visible para los que utilizan un determinado navegador, sea cual sea, está contribuyendo a hacer que la web sea un sitio más fragmentado y peor.

La web tendría que ser universal y para todo el mundo, pero si no seguimos una serie de normas, no todos podrán tener acceso a ella.

Ejemplo

Un ejemplo muy claro es el de los invidentes, que navegan con un sintetizador de voz, que toma el texto de una web y lo lee en voz alta. Por lo tanto, en el caso de las imágenes, por ejemplo, si éstas no contienen una buena descripción de lo que se ve, habrá gente que no podrá acceder a ese contenido.

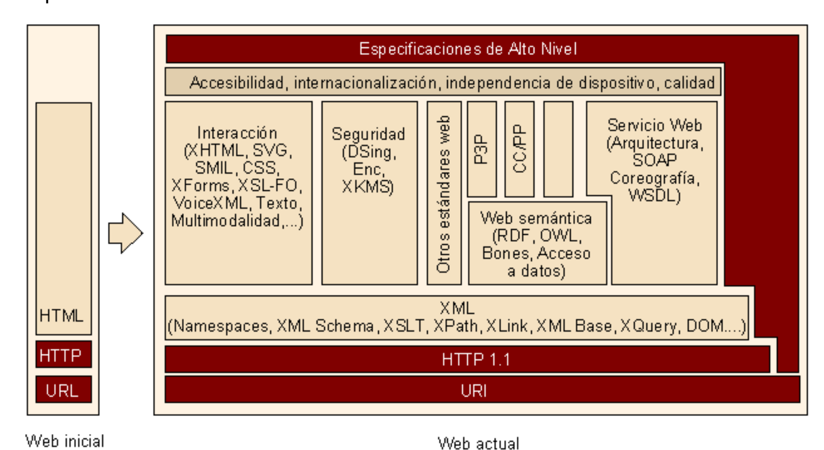
El esfuerzo del W3C ha dado pie a más de **70 tecnologías**. Partieron del HTML para llegar al modelo "actual", donde nos encontramos nuevas tecnologías para modelizar hacia dónde está yendo la web. Nuevas tecnologías para:

- web semántica: RDF, OWL
- servicios web: SOAP y WSDL
- gráficos y multimedia (SVG, SMIL)
- diálogos de voz (VoiceXML)
- formularios interactivos (XForms)
- documentos de texto (XHTML, MathML)
- presentación de contenidos (CSS)
- etc.

La mayoría son aplicaciones del XML, pero también hay tecnologías no basadas en XML (WebCGM, PNG).

Se ha de decir que, mientras que algunas de estas tecnologías ya están bastante maduras, todavía hay otras en pleno desarrollo.

Esquematización de la evolución de los estándares web: del HTML a la web actual



Más información sobre estándares web

Podréis encontrar más información sobre estándares web en Web Standards Project y World Wide Web Consortium W3C.

HTML, XML y XHTML

HTML es el lenguaje de marcaje predominante para la creación de páginas web. Se utiliza para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos como imágenes. El HTML se escribe en forma de etiquetas entre corchetes angulares (<,>). El HTML también puede describir, hasta cierto punto, la apariencia de un documento, y puede incluir *scripts* (por ejemplo, Javascript), lo cual puede afectar al comportamiento de navegadores web y otros procesadores de HTML.

HTML

HTML es el acrónimo inglés de *hypertext markup language*('lenguaje de marcaje de hipertexto').

Por convención, los archivos en formato HTML utilizan la extensión .htm o .html.

El HTML se desarrolló como un subconjunto del SGML (un lenguaje de marcas más complejo) y es el lenguaje de marcas que ha hecho posible la web tal como la conocemos actualmente.

El HTML 4.01 es la última versión de HTML recomendada por el W3C.

Como ya hemos visto, el HTML se definió dentro del marco de SGML y fue de lejos la aplicación más conocida de este estándar. De todos modos, los navegadores nunca han puesto muchas exigencias al código HTML que interpretan, y las páginas web pueden llegar a ser muy caóticas si no cumplen la sintaxis. Sin embargo, hay que decir que no todos los navegadores siguen los estándares, lo cual significa que una misma web se puede visualizar de manera diferente según qué navegador la interprete.

XML es un metalenguaje extensible de etiquetas desarrollado por el World Wide Web Consortium (W3C). Es una simplificación y adaptación del SGML y permite definir la gramática de lenguajes específicos (de la misma manera que HTML es un lenguaje definido por SGML). Por lo tanto, el XML no es realmente un lenguaje en particular, sino una manera de definir lenguajes para diferentes necesidades.

XML

XML es el acrónimo inglés de *extensible markup language* (lenguaje extensible de marcas).

El XML no ha nacido sólo para su aplicación en Internet, sino que se propone como un estándar para el intercambio de información entre diferentes plataformas. Se puede utilizar en bases de datos, editores de texto, hojas de cálculo y casi cualquier cosa imaginable.

Es extensible, lo que quiere decir que, una vez diseñado un lenguaje y puesto en producción, se puede extender añadiendo nuevas etiquetas, de manera que los antiguos consumidores de la vieja versión puedan seguir entendiendo el nuevo formato.

Si un tercero decide utilizar un documento creado en XML, es muy fácil entender su estructura y procesarlo de forma automática, pues mejora la compatibilidad entre aplicaciones.

La tecnología XML busca dar solución al problema de expresar información estructurada de la manera más abstracta y reutilizable posible.

XHTML, acrónimo inglés de *extensible hypertext markup language* (lenguaje extensible de marcaje de hipertexto), es el lenguaje de marcaje pensado para sustituir el HTML como estándar para las paginas web. XHTML es la versión XML de HTML, y por lo tanto tiene, básicamente, las mismas funcionalidades, aunque cumple las especificaciones, más estrictas, del XML.

XHTML

XHTML es el acrónimo inglés de *extensible hypertext markup language* (lenguaje extensible de marcaje de hipertexto).

Su objetivo es avanzar en el proyecto del World Wide Web Consortium de conseguir una web semántica, donde la información y la forma de presentarla estén claramente separadas. En este sentido, el XHTML serviría únicamente para transmitir la información que contiene un documento y se dejaría para las hojas de estilos su aspecto y diseño en diferentes medios (ordenadores, PDA, móviles, impresoras).

El XHTML es el sucesor del HTML. Por ello, muchos lo consideran la "versión actual" del HTML, aunque es una recomendación aparte y al mismo tiempo paralela. El W3C continúa recomendando el uso de XHTML 1.1, XHTML 1.0, o HTML 4.01 para publicar en la web.

Los cambios de HTML en la primera generación de XHTML (es decir, XHTML 1.x) son menores, ya que, principalmente, están destinados a conseguir la conformidad con XML.

El cambio más importante es el requisito de que el documento esté bien formado y que todas las etiquetas estén explícitamente cerradas, tal como se requiere en XML.

Como las etiquetas en XML distinguen entre mayúsculas y minúsculas, la recomendación XHTML ha definido todos los nombres de etiqueta en minúsculas. En XHTML, los valores de los atributos tendrán que cerrarse entre comillas (siempre comillas "dobles").

XHTML 1.0

Para ver todas las diferencias entre los lenguajes de marcaje, podéis consultar de forma detallada la recomendación XHTML 1.0 del W3C.

La web semántica, la accesibilidad y la separación de diseño y contenido

En la actualidad, el World Wide Web se basa principalmente en documentos escritos en HTML, un lenguaje de marcas que sirve principalmente para crear hipertexto en Internet. El lenguaje HTML es válido para adecuar el aspecto visual de un documento e incluir objetos multimedia en el texto (imágenes, esquemas de diálogo). Sin embargo, ofrece pocas posibilidades para categorizar y definir los elementos que configuran el texto más allá de las típicas funciones estructurales.

La web semántica se ocupa de resolver estas deficiencias. Para conseguirlo, dispone de tecnologías de descripción de los contenidos que se combinan para aportar descripciones explícitas de los recursos de la web (XML, XHTML, etc.).

Son informaciones adicionales que describen el contenido, el significado, y la relación de los datos. El objetivo es mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos y automatizar al máximo los trabajos de gestión de información.

De esta forma, el contenido queda descrito como los datos de una base de datos accesible por web, o las etiquetas incluidas en el documento. Estas etiquetas permiten que los gestores de contenidos interpreten y realicen procesos inteligentes de captura y tratamiento de la información.

Lo que intentaremos, por lo tanto, es estructurar y marcar el documento de la forma más clarificadora posible. Esto consiste, como veremos más adelante, en separar totalmente el contenido (XHTML) de la forma en que éste se muestra (CSS), para dar el máximo de información y etiquetar de la forma más descriptiva posible nuestro documento.

De alguna manera, sería como decir "eso es una cabecera", "eso es una lista de ingredientes", "ésta es la etiqueta y éste es el valor"...

El hecho de que nuestro contenido esté mejor descrito hará a la vez que éste sea más accesible para la mayoría de usuarios de Internet. Al final, no se trata más que de llamar y marcar las cosas por lo que son.

Contenidos